International Journal of Agricultural Science and Research (IJASR)
ISSN(P): 2250-0057; ISSN(E): 2321-0087
Vol. 5, Issue 6, Dec 2015, 283-288

CONCEPT OF CONDITIONAL REFERENCE PRIOR IN THE STUDY OF GENOTYPIC VARIATION

RATNA RAJ LAXMI[1], SANDHYA[2] & CHETAN[3]

Department of Statistics, M. D. University, Rohtak, Haryana, India

ABSTRACT

As genetic diversity is a valuable tool for plant breeding and screening promising genotypes. The increase in population and the subsequent rise in the demand for agricultural produce are expected to be greater in regions where production is already insufficient, in particular in South Africa and South Asia. The necessary increase in agricultural production represents a huge challenge to local farming systems and must come mainly from increased yield per unit area. To meet this requirement various crop improvement programmers have been initiated around the globe. Under any crop improvement programme a sample of promising genotypes are performance tested every year. To obtain good samples of genotypes many researchers studying the genotypic variation between/ within genotypes. In this addition to this chain an attempt is to make to evaluate the loss of information during the study of these genotypes using the conditional reference prior. It is observed that that an integrated likelihood approach for studying genotypic variation obtained by using a conditional reference prior is equal to the marginal likelihood approach for parameter obtained by the non-centrality parameter.

KEYWORDS: Genotypic Variation, Conditional Reference Prior, Non-Centrality Parameter, Marginal Likelihood Etc

Original Article

Received: Oct 13, 2015; **Accepted:** Nov 24, 2015; **Published:** Nov 26, 2015; **Paper Id.:** IJASRDEC201538

INTRODUCTION

As increase in population and the subsequent rise in the demand for agricultural produce are expected to be greater in regions where production is already insufficient, in particular in South Africa and South Asia. The necessary increase in agricultural production represents a huge challenge to local farming systems and must come mainly from increased yield per unit area. To meet this requirement various crop improvement programmers have been initiated around the globe. Under any crop improvement programme a sample of promising genotypes are performance tested every year.As genetic diversity is a valuable tool for plant breeding and screening promising genotypes. Many researchers focused their attention to study genotypic variation over the last five decades years and explode the various concepts relating with genotypes. But the concept related with Bayesian inference for genotypic variation yet not is focused.

Bernardo (1979) initiated the reference prior approach to development of non-informative priors. Because of their importance scores of applications and reference prior methods have been progressively defined and refined over the subsequent years. Many research articles recorded the evolutions in the methods including Berger and Bernardo (1989, 1992a, 1992b), Berger, Bernardo and Mendoa (1989) and Ye (1990). Due to extensive range of applicability, an attempt is made to study the concept of conditional reference prior of non-central parameter in genotypic variation. It is observed that, in case of marginal experiments, which provides the quantity F for

genotypic variation, discarding all the information about the single mean genotypes. It is then apparent that there is a loss of information. Here an attempt is made to clarify such loss of information. It has also been shown that the conditional reference prior for the parameter ζ obtain from the non-central parameter under the analysis of genotypic variation has been same when it is obtain through marginal likelihood approach and integrated likelihood approach.

Marginal Likelihood Approach

Let y_{ij}be the observation relative to the j^{th}replication of i^{th} genotype. The terms $\mu_1, \mu_2, \dots, \mu_k$ denote unknown parameters and $\in_{ij}$ represents the random error component, which we suppose to be independent and identically distributed (IID) according to the low $N(0, \sigma^2)$ with unknown. Therefore

$$y_{ij} = \mu_i + \varepsilon_{ij} \ (i = 1,2, \dots K, j = 1,2, \dots r_i)$$

and Null hypothesis $H_0: \mu_1 = \mu_2 =, \dots, = \mu_k = \mu$ (Say)

Against Alternative hypothesis $H_1: \mu_i \neq \mu$

based on the test statistic

$$F = \frac{\sum_i r_i (\bar{y}_i - \bar{y})^2 / u_1}{\sum_i \sum_j (y_{ij} - \bar{y}_i)^2 / u_2} \ u_1 = k - 1, u_2 = N - k$$

Where$\bar{y}_{ij} = \frac{1}{r_i} \sum_j y_{ij}, \bar{y} = \frac{1}{N} \sum_i \sum_j y_{ij}$ and $N = \sum_i r_i$

It is well-known that under the null hypothesis H_0, the $F_{cal.}$ follows a $F_{u_1, u_2, \lambda}$ distribution. Under the alternative hypothesis, the follows non-central distribution, whose non-centrality parameter isλi.e.

$$\lambda = \frac{N}{\sigma^2} \sum_i w_i \left(\mu_i - \bar{\mu}\right)^2$$

Where$\bar{\mu} = \sum_i w_i \mu_i$ and $w_i = \frac{r_i}{N}$

The above testing procedure is a clear and classical example of how the problem of eliminating nuisance parameters is handled when the source of variation are genotypes. Although the model has $k + 1$ parameters, a scalar test statistic is contracted to compare different values of the scalar non-centrality parameter λ. In a certain sense, the classical test acts as if we would have observed only the marginal experiment, which provides the quantity F for genotypic variation, discarding all the information about the single mean genotypes $(\bar{y}_1, \bar{y}_2, \dots, \bar{y}_k)$.

The great simplicity of the standard one way classification on transformation of multi-parametric problem into an equivalent scalar one, where the only parameter involved is the non-centrality parameter, therefore we can rewrite our hypothesis as

$$H_0: \lambda = 0 \text{ against } H_1: \lambda > 0$$

Then, frequency-based solutions to the problem consist only in the computation of sample realization $F_{cal.}$ of the F statistic. Bertolino et al. (1990) propose to consider as the actual result of a marginal experiment and then to use the sampling distribution of the F statistic as a marginal likelihood. They also suggest that it is more convenient to work with

the parameter $\zeta = \frac{\lambda}{N}$ instead of when $\lambda = 0(\text{under} H_0)$ the distribution reduces to the well-known Fisher-Snedecor law. Therefore the marginal likelihood function associated to$F_{cal.}$ is given by

$$L(\zeta) \propto \sum_{j=0}^{\infty} p_j \left(\frac{N\zeta}{2}\right) \frac{\Gamma\left(\frac{u_1+u_2}{2}+j\right)}{\Gamma\left(\frac{u_1}{2}+j\right)} \left(\frac{u_1 F_{cal}}{u_2+u_1 F_{cal.}}\right)^j$$

$$\text{where } p_j(\zeta) = \exp {(-z)z^j}/{j!}$$

If $M(a, b; z)$ is the Kummer's confluent Hypergeometric function, then $L(\zeta)$ is given by

$$L(\zeta) \propto \sum_{j=0}^{\infty} p_j \left(\frac{N\zeta}{2}\right) \left(\frac{u_1+u_2}{2}\right)_j \left(\frac{u_2}{2}\right)_j^{-1} \left(\frac{u_1 F_{cal.}}{u_2+u_{1,} F_{cal}}\right)$$

$$\propto \exp\left\{\frac{-N\zeta}{2}\right\} M \left\{\frac{u_1+u_2}{2}, \frac{u_2}{2}, \frac{n\zeta}{2}, \left(\frac{u_1 F_{cal.}}{u_2+u_{1,} F_{cal}}\right)\right\}$$

Where $(a)_j = \frac{\Gamma(a+j)}{\Gamma(a)}$ is the Pochhammer'ssymbol.

By using the asymptotic approximation for $L(\zeta)$ the non informative prior for the parameter ζ is [Solari(2002)]

$$\pi(\zeta) \propto \frac{1}{\sqrt{\zeta}}$$

Integrated Likelihood Approach

Let $1_p$and 0_p denote the $p \times 1$vector of 1's and 0's resp., then the model can be rewrite as

$$\mathbf{y} = \mathbf{A}\boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Where the **A** matrix is order of $n \times k$ and given by

$$\mathbf{A} = \begin{bmatrix} \mathbf{1}_{N_1} & \mathbf{0}_{N_1} & \cdots & \mathbf{0}_{N_1} \\ & \vdots & \ddots & \vdots \\ \mathbf{0}_{N_k} & \mathbf{0}_{N_k} & \cdots & \mathbf{0}_{N_k} \end{bmatrix}$$

And ϵ is $\mathbf{N}(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I_n})$**.** Then, the parameter $\boldsymbol{\zeta}$can be rewrite as

$$\boldsymbol{\zeta} = \frac{\mathbf{1}}{\boldsymbol{\sigma}^2}(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}\mathbf{1}_k)^1 \mathbf{R}(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}\mathbf{1}_k) = \frac{\mathbf{1}}{\boldsymbol{\sigma}^2}\boldsymbol{\mu}'\mathbf{D}'\mathbf{R}\mathbf{D}\boldsymbol{\mu}$$

where$\mathbf{R} = \frac{\mathbf{1}}{\mathbf{N}}\mathbf{diag}(\mathbf{N}_1, \mathbf{N}_2, \ldots, \mathbf{N}_k)$

$$\mathbf{D} = \frac{\mathbf{1}}{\mathbf{N}} \begin{bmatrix} \mathbf{N} - \mathbf{N}_1 & -\mathbf{N}_2 & \cdots & -\mathbf{N}_k \\ -\mathbf{N}_1 & \mathbf{N} - \mathbf{N}_2 & & -\mathbf{N}_k \\ \vdots & & \ddots & \vdots \\ -\mathbf{N}_1 & \mathbf{N}_2 & \cdots & \mathbf{N} - \mathbf{N}_k \end{bmatrix}$$

Also we can express the parameter of interest (ζ) as

$$\boldsymbol{\zeta} = \frac{\mathbf{1}}{\mathbf{N}\boldsymbol{\sigma}^2}\sum_{\mathbf{i}} \mathbf{N_i}\boldsymbol{\mu}_\mathbf{i}(\boldsymbol{\mu}_\mathbf{i} - \bar{\boldsymbol{\mu}})$$

Or

$$\boldsymbol{\zeta} = \frac{1}{\sigma^2}\boldsymbol{\mu}'\mathbf{R}(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}\mathbf{1_k}) = \frac{1}{\sigma^2}\boldsymbol{\mu}'\mathbf{RD}\boldsymbol{\mu}$$

Since the k-dimensional vector $(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}\mathbf{1_k})$ lies in a (k-1) dimensional subspace of R^k, so we need to transform $(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}\mathbf{1_k})$ into a (k-1)-dimensional vector$\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \dots, \boldsymbol{\theta_{k-1}})'$. As rank (RD) =rank (D) =k $-$ 1, there exists a (k $-$ 1)$\times$ k matrix **P**such that$\mathbf{PP}' = \mathbf{RD}$. Then the Jacobian matrix $\mathbf{J_1}$of the transformation from $(\boldsymbol{\mu}, \boldsymbol{\sigma^2})$ to $(\mathbf{0}, \boldsymbol{\sigma^2}, \bar{\boldsymbol{\mu}})$ is

$$\mathbf{J_1} = \left[\frac{\partial(\boldsymbol{\mu},\boldsymbol{\sigma}^2)}{\partial(\boldsymbol{\theta},\boldsymbol{\sigma}^2,\bar{\boldsymbol{\mu}})}\right] = \begin{bmatrix} \mathbf{PR^{-1}} & \mathbf{0_{k-1}} \\ \mathbf{0'_k} & \mathbf{1} \\ \mathbf{1'_k} & \mathbf{0} \end{bmatrix}$$

Where $\boldsymbol{\mu} = \mathbf{R^{-1}P'}\boldsymbol{\theta} + \bar{\boldsymbol{\mu}}\mathbf{1_k}$

Now we can reparameterizeθ in the new parameter

$$\boldsymbol{\eta} = \boldsymbol{\theta}'\boldsymbol{\theta} = \frac{1}{N}\sum_i \mathbf{N_i}(\boldsymbol{\mu_i} - \bar{\boldsymbol{\mu}})^2$$

and

$$\boldsymbol{\psi} = (\boldsymbol{\psi_1}, \boldsymbol{\psi_2}, \dots, \boldsymbol{\psi_{k-2}})\boldsymbol{\epsilon}$$

$\boldsymbol{\psi} = (0, \pi)^{k-3} \times (0{,}2\pi)$[Berger et al. (1998)]. Then

$$\boldsymbol{J_2} = \left[\frac{\partial(\mathbf{0},\boldsymbol{\sigma}^2,\bar{\boldsymbol{\mu}})}{\partial(\boldsymbol{\eta},\boldsymbol{\sigma}^2,\bar{\boldsymbol{\mu}},\boldsymbol{\psi})}\right] = \begin{bmatrix} \boldsymbol{\eta}^{1/2}\boldsymbol{b}' & \mathbf{0} & \mathbf{0} \\ \mathbf{0}'_{\boldsymbol{k-1}} & \mathbf{1} & \mathbf{0} \\ \mathbf{0}'_{\boldsymbol{k-1}} & \mathbf{0} & \mathbf{1} \\ \boldsymbol{\eta}^{1/2}\boldsymbol{c} & \mathbf{0}_{\boldsymbol{k_2}} & \mathbf{0}_{\boldsymbol{k_2}} \end{bmatrix}$$

Where the $(k-1)$-dimensional vector **b** and the $(k-2)\times(k-1)$ matrix $\times$**c** are given by

$$\boldsymbol{b} = \boldsymbol{\eta}^{1/2}\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{\eta}}$$

$$\boldsymbol{c} = \boldsymbol{\eta}^{-1/2}\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{\psi}}$$

Since $\zeta = \frac{\eta}{\sigma^2}$, the Jacobian matrix $\boldsymbol{J_3}$ of the transformation from $(\boldsymbol{\eta}, \boldsymbol{\sigma^2}, \bar{\boldsymbol{\mu}}, \boldsymbol{\psi})$ to $(\boldsymbol{\zeta}, \boldsymbol{\sigma^2}, \bar{\boldsymbol{\mu}}, \boldsymbol{\psi})$ is given

$$\boldsymbol{J_3} = \left[\frac{\partial(\boldsymbol{\eta},\boldsymbol{\sigma}^2,\bar{\boldsymbol{\mu}},\boldsymbol{\psi})}{\partial(\boldsymbol{\zeta},\boldsymbol{\sigma}^2,\bar{\boldsymbol{\mu}},\boldsymbol{\psi})}\right] = \begin{bmatrix} \boldsymbol{\sigma^2} & \mathbf{0} & \mathbf{0} & \mathbf{0}'_{\boldsymbol{k_2}} \\ \boldsymbol{\zeta} & \mathbf{1} & \mathbf{0} & \mathbf{0}'_{\boldsymbol{k-2}} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0}'_{\boldsymbol{k-2}} \\ \mathbf{0}_{\boldsymbol{k-2}} & \mathbf{0}_{\boldsymbol{k-2}} & \mathbf{0}_{\boldsymbol{k-2}} & \boldsymbol{I}_{\boldsymbol{k-2}} \end{bmatrix}$$

Here we defined a new set of parameters $(\zeta, \sigma^2, \bar{\mu}, \psi)$ in which ζ is the parameters of interest and $\xi = (\sigma^2 . \bar{\mu}, \psi)$ is the nuisance parameters. For this new parameters (ζ, ξ) is given by

$$I(\zeta, \xi) = (\boldsymbol{J_3}, \boldsymbol{J_2}, \boldsymbol{J_1})\boldsymbol{I}(\boldsymbol{\mu}, \boldsymbol{\sigma^2})(\boldsymbol{J_1}, \boldsymbol{J_2}, \boldsymbol{J_3})'$$

Where $\boldsymbol{I}(\boldsymbol{\mu}, \boldsymbol{\sigma^2}) = \begin{bmatrix} \frac{1}{\sigma^2}\boldsymbol{A}'\boldsymbol{A} & \mathbf{0}_{\boldsymbol{k}} \\ \mathbf{0}'_{\boldsymbol{k}} & \frac{N}{2(\sigma^2)^2} \end{bmatrix} = \boldsymbol{N}\begin{bmatrix} \frac{1}{\sigma^2}\boldsymbol{R} & \mathbf{0}_{\boldsymbol{k}} \\ \mathbf{0}'_{\boldsymbol{k}} & \frac{1}{2(\sigma^2)^2} \end{bmatrix}$

$$\boldsymbol{I}(\boldsymbol{\zeta},\boldsymbol{\xi}) = Ndiag\left(\begin{bmatrix}\frac{b'b}{\zeta} & \frac{b'b}{\sigma^2}\\ \frac{b'b}{\sigma^2} & \frac{\zeta b'b+\frac{1}{2}}{(\sigma^2)^2}\end{bmatrix}, \frac{1}{\sigma^2}, \boldsymbol{\zeta cc'}\right)$$

$$= Ndiag\left(\begin{bmatrix}\frac{1}{4\zeta} & \frac{1}{4\sigma^2}\\ \frac{1}{4\sigma^2} & \frac{\zeta+2}{4(\sigma^2)^2}\end{bmatrix}, \frac{1}{\sigma^2}, \boldsymbol{\zeta cc'}\right)$$

Because b’b = $\frac{1}{4}$

$$\equiv Ndiag\left(\begin{bmatrix}\boldsymbol{i}_{11} & \boldsymbol{i}_{12}\\ \boldsymbol{i}_{21} & \boldsymbol{i}_{22}\end{bmatrix}, \boldsymbol{i}_{33}, \boldsymbol{I}_{44}\right)\text{(Say)}$$

Now follow the reference prior algorithm [Berger and Bernardo 1992], the conditional reference priors for (ζ, ξ) are given by

$$\boldsymbol{\pi}(^{\boldsymbol{\psi}}/_{\boldsymbol{\xi}}) \propto |\boldsymbol{I}_{44}|^{1/2} \propto \prod_{i=1}^{k-3}(\boldsymbol{sin\psi_i})^{k-i-2}$$

$$\boldsymbol{\pi}(^{\overline{\boldsymbol{\mu}}}/_{\boldsymbol{\zeta}}, \boldsymbol{\sigma}^2) \propto |\boldsymbol{i}_{33}|^{\frac{1}{2}} \propto \mathbf{1}$$

$$\boldsymbol{\pi}(^{\boldsymbol{\sigma}^2}/_{\boldsymbol{\zeta}}) \propto |\boldsymbol{i}_{22}|^{1/2} \propto \frac{1}{\sigma^2}$$

$$\boldsymbol{\pi}(\boldsymbol{\zeta}) \propto |\boldsymbol{i}_{22}|^{-1/2}|\boldsymbol{i}_{11}\boldsymbol{i}_{22} - \boldsymbol{i}_{12}\boldsymbol{i}_{21}|^{1/2} \propto \frac{1}{\sqrt{\zeta(\zeta+2)}}$$

If we constructed the reverse reference prior, where the reverse ordering of the parameter $(\psi, \bar{\mu}, \sigma^2, \zeta)$ or (ξ^*, ζ) the prior distribution of the parameter of interest ζ conditionally on the nuisance parameter ξ^* is given by

$$\boldsymbol{I}(\boldsymbol{\xi}^*, \boldsymbol{\zeta}) = Ndiag\left(\boldsymbol{I}_{44}, \boldsymbol{i}_{33}, \begin{bmatrix}\boldsymbol{i}_{11} & \boldsymbol{i}_{12}\\ \boldsymbol{i}_{21} & \boldsymbol{i}_{22}\end{bmatrix}\right)$$

$$= Ndiag\left(\boldsymbol{\zeta cc'}, \frac{1}{\sigma^2}, \begin{bmatrix}\frac{\zeta+2}{4(\sigma^2)^2} & \frac{1}{4\sigma^2}\\ \frac{1}{4\sigma^2} & \frac{1}{4\zeta}\end{bmatrix}\right)$$

Therefore conditional reference priors are

$$\boldsymbol{\pi}^*\left(^{\boldsymbol{\zeta}}/_{\boldsymbol{\xi}^*}\right) \propto |\mathbf{i}_{11}|^{1/2} \propto \frac{1}{\sqrt{\zeta}}$$

$$\boldsymbol{\pi}^*\left(^{\boldsymbol{\sigma}^2}/_{\boldsymbol{\psi}}, \overline{\boldsymbol{\mu}}\right) \propto |\mathbf{i}_{22}|^{\frac{1}{2}} \propto \frac{1}{\sigma^2}$$

$$\boldsymbol{\pi}^*\left(^{\overline{\boldsymbol{\mu}}}/_{\boldsymbol{\xi}}\right) \propto |\mathbf{i}_{33}|^{1/2} \propto \mathbf{1}$$

$$\boldsymbol{\pi}^*(\boldsymbol{\psi}) \propto |\mathbf{I}_{44}|^{1/2} \propto \prod_{i=1}^{k-3}(\sin\boldsymbol{\psi}_i)^{k-i-2}$$

CONCLUSIONS

This shows that an integrated likelihood approach for studying genotypic variation obtained by using a conditional reference prior is equal to the marginal likelihood approach for parameter ζ. It is also found that when ζ is the real quantity of interest the marginal likelihood is a correct report of the information available.

REFERENCES

1. *Berger, J. and Bernardo, J.M. (1989). Estimating a product of mean: Bayesian analysis with reference priors, Journal of American Statistics Association, 84, 200-207.*
2. *Berger, J. and Bernardo, J.M. (1991a). Reference priors in a variance components problem. Proceeding of Indo-USA Workshop on Bayesian Analysis in Statistics and Econometrics, P.Goel (ed.), Bangalore, India.*
3. *Berger, J. and Bernardo, J.M. (1991b). Ordered group reference prior with application to a multinomial problem, Biometrika, 79, 25-37.*
4. *Berger, J. O., Philippe, A., Robert, C. P. (1998). Estimation of quadratic functions: noninformative priors for non-centrality parameters. Statistica Sinica 8, 359–375.*
5. *Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. Journal of the Royal Statistical Society B 41, 113–147 (with discussion).*
6. *Bertolino, F., Piccinato, L., Racugno, W. (1990). A marginal likelihood approach to analysis of variance. The Statistician 39, 415–424.*
7. *Solari, F. (2002). L'uso di pseudo-verosimiglianzenell'analisidellavarianza. Phd thesis, Dipartimento di statistica, probabilità e statisticheapplicate, Università di Roma Sapienza.*
8. *Ye, K.Y. (1990). Noninformative prior in Bayesian analysis, Ph.D. Thesis, Department of Statistics, Purdue University, W. Lafayette.*